ANNEXURE

Course Name: Certificate Course in Data Analysis & Machine Learning using python

Course Objective: This course provides a comprehensive journey from fundamental data manipulation to advanced predictive modeling using the Python ecosystem. Participants will master core libraries like Pandas for data wrangling, Seaborn for visualization, and Scikit-learn for building robust machine learning pipelines. The curriculum focuses on transforming raw business data into actionable intelligence by balancing statistical foundations with hands-on algorithmic implementation.

Prerequisite: Candidates should be proficient in Computer Fundamentals, Basic Database and Programming Concepts with Logical Approach.

Course Duration: 80 Hrs (8 hours/ day for 2 Weeks)

Teaching Schema

S. No.	Modules	Hours
1	Introduction to Python programming language	8
2	Understanding the concept of data structures in Python and their implementation	8
3	Pandas basics for data manipulation in Python	8
4	Data visualization on using matplotlib and seaborn libraries	8
6	Machine Learning Foundations	8
7	Supervised Learning - Regression	8
8	Supervised Learning - Classification	8
9	Unsupervised Learning & Model Refinement	8
10	Project	8
	Total	**80**

Detailed Course Content

1. **Introduction to Python programming language:**
    - Installation and setup,
    - Creating and saving a script file
    - basics of Python,
    - data types,
    - Variables.
    - Syntax and comments in Python: Understanding the syntax and commenting the code
    - Python string manipulation: String data type, string indexing, slicing, concatenation, and formatting
    - Introduction to Python operators:
        - Arithmetic operators,
        - comparison operators,
        - logical operators,
        - Assignment operators.
2. **Understanding the concept of data structures in Python and their implementation**
    - Python lists,
    - tuples,
    - sets,

- dictionaries

Conditional statements and loops in Python:

- if-else,
- while loop,
- for loop,
- Nested loops.

Introduction to Python functions:

- Defining functions, arguments, return statement, and scope of variables.

Introduction to Data Manipulation

- Introduction to NumPy

- NumPy Package in Python, Importing NumPy, creating different arrays using NumPy, Array Functions and Methods, Different Mathematical Functions, Different Matrix Operations, Random Numbers, Generate Numbers between a ranges.

3. Pandas basics for data manipulation in Python:

- Understanding the Pandas library, series, and data frame operations.
- Concept of Series in Pandas, Creating Series using Pandas, Different Series Attributes, Series vs List, Series Operations, Series from CSV File.

Exploratory data analysis,

- Reading files
- Data cleaning in Python: handling missing values and filling NA
- Data preparation and pre-processing
- Data feature engineering: handling categorical data
- Data validation techniques in Python
- Data feature engineering: removing columns and rows from raw data

4. Data visualization on using matplotlib and seaborn libraries

- Scatter plot
- Line plot
- Bar plot
- Histogram
- Box plot
- Pie plot

5. Machine Learning Foundations (8 Hours)

- AI vs. ML vs. DL: Clarifying the hierarchy and types of learning (Supervised vs. Unsupervised).
- Problem definition, data prep, model training, and deployment.
- Pre-processing for AI: Feature scaling (Standardization/Normalization) and Label Encoding.

6. Supervised Learning - Regression (10 Hours)

- Simple and Multiple regression models.
- Understanding coefficients, intercepts, and the "Ordinary Least Squares" method.
- Using R-Squared, Mean Absolute Error (MAE), and RMSE.
- Predicting real estate prices or sales forecasting.

7. Supervised Learning - Classification (10 Hours)

- Logistic Regression: Probability-based classification.
- Decision Trees, Random Forests, and K-Nearest Neighbors (KNN).
- The Confusion Matrix: Understanding Accuracy, Precision, Recall, and F1-Score.
- **Project:** Predicting customer churn or credit risk assessment.

8. Unsupervised Learning & Model Refinement (12 Hours)

- Clustering K-Means and Hierarchical clustering for customer segmentation.
- Dimensionality Reduction: Introduction to PCA (Principal Component Analysis).
- Model Optimization: Hyper parameter tuning using GridSearchCV and Cross-Validation.

9. **Final Project:** showcase all your skills in an end-to-end data analysis project. You'll pick the dataset, do the data munging, ask the research questions, visualize the data, draw conclusions, and present your result